AFGL-TR-83-0040

ON THE VARIANCE OF PROBABILITY ESTIMATES WITH CORRELATED DATA

Robert R. Beland

Rio Grande Associates
921 Teetshorn
Houston, TX 77009

Scientific Report No. 1

November 1982

AIR FORCE GEOPHYSICS LABORATORY
AIR FORCE SYSTEMS COMMAND
UNITED STATES AIR FORCE
HANSCOM AFB, MASSACHUSETTS 01731

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE	READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER: AFGL-TR-83-0040 — 2. GOVT ACCESSION NO.: AD-A126 336	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle): On the Variance of Probability Estimates with Correlated Data	5. TYPE OF REPORT & PERIOD COVERED: Scientific Report #1
	6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s): Robert R. Beland	8. CONTRACT OR GRANT NUMBER(s): F19628-81-C-0128
9. PERFORMING ORGANIZATION NAME AND ADDRESS: RIO GRANDE ASSOCIATES, 921 Teetshorn, Houston, TX 77009	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: 62101F, 9993XXXX
11. CONTROLLING OFFICE NAME AND ADDRESS: Air Force Geophysics Laboratiry, Hanscom AFB, MA 01731, Monitor: Paul Tsipouras/SUNA	12. REPORT DATE: November 1982
	13. NUMBER OF PAGES: 30
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)	15. SECURITY CLASS. (of this report): Unclassified
	15a. DECLASSIFICATION DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

correlated data: stochastic processes: empirical distribution function: amplitude histogram: Gaussian processes; variance of probability estimates: probability estimation.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The foundations for probability estimation are presented for random processes in discrete time. The estimators considered are the empirical distribution function and the amplitude histogram and conditions are given for these to be unbiased and consistent. Expressions for the variance and covariance of these estimators are presented and the effect of correlations on these quantities is discussed and demonstrated numerically by simulations. A theorem establishes a monotonic relation between the variance and the correlations for Gaussian processes.

DD FORM 1 JAN 73 1473

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

TABLE OF CONTENTS

Page

LIST OF TABLES

ABSTRACT

The foundations for non-parametric probability estimation are presented for random processes in discrete time. The estimators considered are the empirical distribution function and the amplitude histogram. It is shown that if the process is strictly stationary and satisfies a mixing condition, the estimators are unbiased and consistent. Expressions for the variance and covariance of these estimators are presented and the effect of the correlations on these quantities is discussed. This effect is demonstrated numerically by simulations of a Gaussian process. A theorem is established which demonstrates the monotonic relation between the variance and the correlations for Gaussian processes. This follows from a corresponding property of the bivariate Gaussian distribution.

1. INTRODUCTION

A common problem in many areas of science is the treatment of correlated data. Classical statistical methods which assume independence of observations (and this is the great body of statistics) are generally not applicable. This research addresses one aspect of this larger problem, that of probability estimation with correlated data.

This problem arises in many diverse fields where models based on time series or stochastic processes are used. In some of these areas, a non-parametric estimate of the univariate probability is of intrinsic interest. An area of special importance is that where the model is a stationary Gaussian random process. This importance derives from the parametric simplicity of Gaussian processes and from the fact that many techniques assume or require Gaussianity for their validity. This is especially true of many signal processing techniques. For example, Gaussianity is usually assumed in the smoothing, prediction and signal extraction problems and is required for the equivalence of maximum likelihood and least squares methods of estimation. Given the current popularity of maximum entropy methods (MEM) of spectral estimation, it is worth pointing out that such estimates maximize the entropy only for Gaussian processes.

Thus, there is ample motivation for the characterization of probability estimation based on finite samples of random processes. Although there has been exhaustive work in such estimation for independent variables (see., e.g., Wegman [1972]), little has been done for random processes. Patankar (1954) and Thrall (1965) seem to be the only authors to have addressed this problem. The related problem of hypothesis testing for the Gaussianity of a stochastic process has received more attention (Persson [1974], Gasser [1975], Weiss [1978]). Most of this latter work has been motivated by the particular statistical requirements involved in the analysis of the electroencephalogram. (See also McEwen and Anderson [1975], Saunders [1963], Elul [1969].) Since most of these authors use test statistics constructed from univariate distribution estimates (i.e., the

Kolmogorov-Smirnov or Chi-squared tests), it is surprising the foundation for this distribution estimation has not been more thoroughly explored.

In this work, we will investigate the estimation of the univariate distribution function based on a finite sample of a stochastic process. We will restrict our attention to strictly stationary random processes defined on discrete time. The estimators that will be treated are the empirical distribution function and the amplitude histogram. We will first establish conditions for these estimators to be unbiased and consistent. The variance and covariance of these estimators will be presented and it is here that the presence of correlations manifests itself most strikingly. The effect of correlations on these variances will be investigated in more detail for the particular case of Gaussian processes. Simulations of a Gaussian process wil be performed to numerically demonstrate the effect of correlations. These numerical results will then be generalized by a theorem which shows that, for Gaussian processes, these variances are monotonically related to the correlations.

2. SOME PROPERTIES OF THE ESTIMATORS

In this section, we will develop some of the statistical properties of the estimators required by the Kolmogorov-Smirnov and Chi-squared tests. These estimators are, respectively, the empirical distribution function and the so-called amplitude histogram or distribution. These are closely related. Parzen (1962a) has established these properties and pursued them into the realm of non-parametric probability estimation for the case of independent, identically distributed variables. Thrall (1965) has published some of the following results for the case when correlations are present, but also dealt primarily with independent observations.

Suppose we have n successive observations of a stationary stochastic process X(t) defined on the integers. We can assume, by stationarity, that the observations are for times 1,2.....,n. We form the sample or empirical distribution function, $F_n(a)$ as follows;

$$F_n(a) = \frac{1}{n} \sum_{i=1}^{n} I_a(X(i))$$

where $I_a(.)$ is the indicator function of the set $[-\infty, a]$ defined by:

$$I_a(x) = \begin{cases} 1 \text{ if } x \leq a \\ 0 \text{ if } x > a. \end{cases}$$

It is trivial to show that stationarity implies that

$$E[F_n(a)] = F(a)$$

where F(.) is the univariate distribution. It should be emphasized that this unbiasedness holds even in the presence of correlations or dependencies between variables.

Next, we wish to evaluate variances. We have:

$$\mathrm{Var}[F_n(a)] = \mathrm{Var}[\frac{1}{n}\sum_{1}^{n} I_a(X(i))]$$

$$= \frac{1}{n^2}\{\sum_{1}^{n} \mathrm{Var}[I_a(X(i))]$$

$$+\sum_{1}^{n}\sum_{=j}^{n} \mathrm{Cov}[I_a(X(i)),I_a(X(j))]\}.$$

But we also have:

$$\mathrm{Var}[I_a(X(i))] = F(a)[1 - F(a)]$$

Upon substitution, we get:

$$\mathrm{Var}[F_n(a)] = \frac{1}{n} F(a)[1-F(a)]$$

$$+ \frac{1}{n^2} \sum_{i}^{n}\sum_{\neq j}^{n} \mathrm{Cov}[I_a(X(i)),I_a(X(j))].$$

Note that this expression depends on the covariance of the random variable $I_a(X(i))$, which is a binary (i.e., 0 or 1) stochastic process. If the variables X(1) and X(j) are independent, then the covariance vanishes and the variance reduces to:

$$\mathrm{Var}[F_n(a)] = \frac{1}{n} F(a)[1-F(a)].$$

This expression is easily understood because, in the case of independence, $I_a(X(i))$ is a simple Bernoulli variable with parameters p=F(a) and q=[1-F(a)]. In the more general case, our variance is in the form of the sum of two terms: the first is for independent variables and the second is the correction for dependence. Because of stationarity, this can be further simplified. Stationarity implies that for all integers h,

$$\mathrm{Cov}[I_a(X(t)),I_a(X(S))] = \mathrm{Cov}[I_a(X(t+h)),I_a(X(s+h))].$$

Thus, the double sum can be transformed and we obtain the same form for the variance as obtained by Thrall (1965):

$$\mathrm{Var}[F_n(a)] = \frac{1}{n} F(a)[1-F(a)]$$

$$+ \frac{2}{n} \sum_{i=1}^{n} (1- \frac{i}{n})\mathrm{Cov}[I_a(X(0),I_a(X(i))].$$

Parzen (1962b) gives the following theorem which states conditions for $F_n(a)$ to be a consistent estimator of $F(a)$.

<u>Theorem 1 (Parzen)</u>: $\mathrm{Var}[F_n(a)]$ converges to zero if and only if $\mathrm{Cov}[F_n(a),I_a(X(n))]$ converges to zero.

This theorem supplies the conditions necessary for a stochastic process to be ergodic for the distribution function, i.e., for the time-averaged sample distribution function (the empirical distribution function) to be equal to the ensemble distribution function. This condition is similar to an α-mixing condition (see, e.g., Billingsley [1979]), which means heuristically that variables that are temporally far apart are effectively independent.

Since we are interested in generalizing the estimation framework developed for independent variables, the notion of α-mixing is especially useful as a generalization of the notion of independence. We next establish the consistency property in terms of this mixing condition.

<u>Corollary 1.1</u>: If the stationary stochastic process $X(t)$ satisfies an α-mixing condition, then $\mathrm{Var}[F_n(a)]$ converges to zero.

Proof: By definition of α-mixing, we have that there exists a sequence α_j such that $0 \leq \alpha_j \leq 1$ for all j and α_j converges to zero and

$$-\alpha_j \leq P(X(0)\leq a),X(j)\leq a) - P(X(0)\leq a)P(X(j)\leq a) \leq \alpha_j$$

for any real a. Thus, after a little algebra with the indices, we get:

$$-\frac{1}{n}\sum_{1}^{n} \alpha_j \leq \mathrm{Cov}[F_n(a),I_a(X(n))] \leq \frac{1}{n}\sum_{1}^{n} \alpha_j.$$

Now, the right and left sides are Cesaro sums of the sequence α_j. If we define C_n by

$$C_n = \frac{1}{n}\sum_{1}^{n} \alpha_j$$

then it is easily shown (see, e.g., Fuller [1976]) that if α_j converges to zero, so does C_n. Thus, we have that:

$$\mathrm{Cov}[F_n(a),I_a(X(n))] \rightarrow 0$$

and so by Theorem 1, $\mathrm{Var}[F_n(a)] \rightarrow 0$.

Q.E.D.

Now let us turn to the covariance of the empirical distribution function. Let a and b be real numbers and assume that a<b. After manipulating the sum and using stationarity, the covariance may be written as:

$$\mathrm{Cov}(F_n(a),F_n(b)) = \frac{1}{n} F(a)[1-F(b)]$$

$$+ \frac{1}{n}\sum_{1}^{n}(1-\frac{i}{n})[P(X(i)\leq a,X(0)\leq b)-F(a)F(b)$$

$$+ P(X(0)\leq a,X(i)\leq b)-F(a)F(b)]$$

The first term is that for the case of independence and the second is the correction for dependence. We also have the following corollary whose proof is omitted, but is very similar to the previous proof.

<u>Corollary 1.2</u>: If the stationary stochastic process X(t) satisfies α-mixing, then $\mathrm{Cov}(F_n(a),F_n(b))$ converges to zero for any a and b.

Let us review our results. We assume that we have n successive observations of a stationary stochastic process defined for discrete time and this process satisfies an α-mixing condition. We have seen that the empirical distribution function is unbiased and consistent as an estimator of the cumulative distribution function. However, the variance and covariance depend strongly on the form of the dependence between variables. The variance is a measure of the average rate of convergence (in mean square) of the estimator $F_n(a)$ to F(a). Thus, we have that the rate of mean square convergence at a point depends on the form of the dependence between variables. This perspective will be pursued in the following section.

Let us next introduce the amplitude histogram. Define the estimator of the amplitude histogram by:

$$f_n(a,b) = F_n(b) - F_n(a)$$

where it is assumed that b>a. We clearly have:

$$E[f_n(a,b)] = F(b) - F(a).$$

The histogram estimator is usually used to construct an estimator of the probability density by dividing it by the interval length (b-a). From the above, we see that it is a biased estimator of the density. The difficulty here stems from the fact that the density is the derivative of the distribution function.

The above bias has led several authors (Parzen [1962a], Leadbetter and Watson [1961]) to use so-called kernel estimates for the density (Wegman [1972]). The problem of bias will be circumvented by considering the histogram as an estimator of a theoretical amplitude histogram, f(a,b), defined by:

$$f(a,b) = F(b) - F(a).$$

The variance of this estimator is easily found from our previous results. It is:

$$\mathrm{Var}(f_n(a,b) = \frac{1}{n} [F(b)-F(a)-(F(B)-F(a))^2$$

$$+ \sum_{i=1}^{n} (1-\frac{i}{n})[P(X(0)\le a,X(i)\le a)-F(a)^2+P(X(0)\le b,X(i)\le b)$$

$$-F(b)^2-P(X(0)\le a,X(i)\le b)-P(X(0)\le b,X(i)\le a)$$

$$+ 2F(a)F(b)]]$$

Note that again we have a term for the independent case and another for the correction due to dependence. The covariance of two such estimators, say $f_n(a,b)$ and $f_n(c,d)$, may also be written down in a straightforward manner, but we will not do so here. We immediately have the following corollary which is a simple extension of earlier results.

Corollary 1.3: If the stationary stochastic process $X(t)$ defined for integer time satisfies α-mixing, then

$$\mathrm{Var}(f_n(a,b)) \rightarrow 0 \text{ for any } a \text{ and } b.$$

We shall end this section with a brief discussion of error limits on the estimator $F_n(a)$. The variance of $F_n(a)$ and the covariance of $F_n(a)$ and $F_n(b)$ are characterized by their dependence on the first and second order distributions of $X(t)$. As noted earlier, we may consider the indicator function as generating a binary process. From this perspective, the variance and covariance depend on the covariance function of the binary process $I_a(X(t))$. In the case of independence, these covariances vanish and the variance of $F_n(a)$ can be estimated in a straightforward fashion from:

$$\mathrm{Var}[F_n(a)] \cong \frac{1}{n} F_n(a)[1 - F_n(a)].$$

For the case of dependence, no such simple procedure is possible. The analogous procedure would require estimates of the second order distribution, which is a computationally formidable task, and one which requires very large amounts of data. However, a variance estimate can be obtained by dealing with the generated binary process $I_a(X(t))$. The covariance function of $I_a(X(t))$ can be estimated using standard techniques. This method is amenable to computational streamlining by the techniques of calculating autocovariances by the Fast Fourier Transform (FFT) algorithm (see, e.g., Oppenheim and Schafer [1975]). The covariance function of $I_a(X(t))$ can then be substituted into the expression for the variance of $F_n(a)$ to yield the desired estimate. Such an approach, although an improvement over direct estimation of the second order distribution, would not be practical due to the complexity and time requirements of the computations.

3. THE VARIANCE OF THE ESTIMATORS FOR GAUSSIAN PROCESSES

In this section, some properties of the variance of the estimators $F_n(a)$ and $f_n(a,b)$ will be explored. We are motivated to treat the simplest case. Gaussian processes are well known for their mathematical tractability. The primary property of Gaussian processes which is important here, indeed, it is indispensable, is the equivalence between the notions of independent and uncorrelated.

Let X_1 and X_2 be random variables that possess the bivariate normal distribution with means 0, variances σ^2, and correlation ρ. Then:

$$P(X_1 \leq a, X_2 \leq b) = \int_{-\infty}^{a} \int_{-\infty}^{b} \frac{1}{2\pi\sigma^2\sqrt{1-\rho^2}} \exp\left[\frac{-1}{2\sigma^2(1-\rho^2)}(x^2+y^2-2\rho xy)\right]dxdy.$$

The double integral cannot be expressed in closed form, i.e., as a function of ρ. It can, however, be evaluated numerically and tabulated (see, e.g., Abramowitz & Stegun [1964]). Some of these values are listed in Table 1 for the cases a=b=0, a=b=0.5, and a=0, b=0.5 These were evaluated numerically using a "double" 8-point Gaussian integration method, that is, the inner integral was evaluated at the Gaussian points (determined by the outer integral) by an 8-point formula. These values allow us to get some numerical estimates for the variance and covariance of the estimator $F_n(a)$ in certain circumstances and so get a qualitative understanding for the effects of correlation.

ρ	$P(X\leq 0, Y\leq 0)$	$P(X\leq 0, Y\leq 0.5)$	$P(X\leq 0.5, Y\leq 0.5)$
-.90	0.0718	0.2031	0.3893
-.80	0.1024	0.2222	0.3981
-.70	0.1266	0.2403	0.4083
-.60	0.1476	0.2572	0.4192
-.50	0.1667	0.2731	0.4305
-.40	0.1845	0.2884	0.4405
-.30	0.2015	0.3031	0.4421
-.20	0.2180	0.3175	0.4539
-.10	0.2341	0.3317	0.4659
0.0	0.2500	0.3457	0.4781
.10	0.2659	0.3598	0.4907
.20	0.2821	0.3740	0.5036
.30	0.2985	0.3884	0.5171
.40	0.3155	0.4031	0.5312
.50	0.3333	0.4183	0.5462
.60	0.3524	0.4343	0.5624
.70	0.3734	0.4512	0.5805
.80	0.3976	0.4693	0.6015
.90	0.4282	0.4884	0.6283

Table 1: Values of bivariate Gaussian probabilities with means 0, variances 1 and correlation ρ .

The only case we can readily compute is the case where the Gaussian process satisfies the equation:

$$X(t) = \frac{1}{1+\theta^2} [a(t) + \theta a(t-1)]$$

where the sequence of variables a(t) are independent, standard normal variables (i.e., $a(t) \sim N(0,1)$). Thus, X(t) has a simple moving average form (Box & Jenkins [1976]) such that:

$$E[X(t)] = 0$$

$$Cov[X(s),X(t)] = \begin{cases} 1 & \text{if } s=t \\ \frac{\theta}{1+\theta^2} & \text{if } |s-t| = 1 \\ 0 & \text{if } |s-t| > 1 \end{cases}$$

Note that the correlation or dependence is only between adjacent values. Note also that the defining equation is normalized so that both X(t) and a(t) have unit variance.

If n is reasonably large, we may write the variance of $F_n(a)$ for this model as:

$$nVar[F_n(a)] = F(a)[1-F(a)] + 2[P(X_0 \le a, X_1 \le a) - F(a)^2]$$

which has a right hand side which is independent of n. To compare with the case of independence, we may calculate the ratio of this variance to the variance for the case of independent variables:

$$R = \frac{F(a)[1-F(a)] + 2[P(X_0 < a, X_1 < a) - F(a)^2]}{F(a)[1-F(a)]}$$

Table 2 exhibits values of $nVar[F_n(a)]$ for correlations in the range (-.5,.5). This range is used because, as is easy to show by standard calculus techniques, the correlation, $\frac{\theta}{1+\theta^2}$, achieves a maximum of 0.5 (θ = 1.0) and a minimum of -0.5 (θ =-1.0) for the first order moving average model. Note that Table 2 uses the values from Table 1. Finally, to accentuate the effect, a column is included which lists the percent change from the independent case.

Tables 1 and 2 exhibit some interesting properties that are the basis for several of the following theorems. First, Table 1 shows that $P(X_1 \le a, X_2 \le b)$ is a monotonically increasing function of the correlation for certain values of a and b. That this property holds for any a and b will be established in Lemma 1. Similarly, Table 2 shows that $nVar[F_n(a)]$, $nVar[F_n(b)]$ and $nCov[F_n(a),F_n(b)]$ also seem to be monotonically increasing functions of the correlation. This result will also be generalized in Theorem 2.

a) ρ	$nVar[F_n(0)]$	R	% change
-0.5	0.08	0.32	-68
-0.4	0.12	0.48	-52
-0.3	0.15	0.60	-40
-0.2	0.19	0.76	-24
-0.1	0.22	0.88	-12
0.0	0.25	1.00	0
0.1	0.28	1.12	12
0.2	0.31	1.24	24
0.3	0.35	1.40	40
0.4	0.37	1.48	48
0.5	0.41	1.64	64

b) ρ	$nVar[F_n(.5)]$	R	% change
-0.5	0.12	0.55	-45
-0.4	0.14	0.65	-35
-0.3	0.15	0.66	-33
-0.2	0.16	0.77	-23
-0.1	0.19	0.89	-11
0.0	0.21	1.00	0
0.1	0.24	1.12	12
0.2	0.26	1.24	24
0.3	0.29	1.37	37
0.4	0.32	1.50	50
0.5	0.35	1.64	64

c) ρ	$nCov[F_n(0),F_n(.5)]$	R	% change
-0.5	0.009	0.02	-94
-0.4	0.04	0.26	-74
-0.3	0.07	0.45	-55
-0.2	0.10	0.63	-36
-0.1	0.13	0.84	-16
0.0	0.15	1.00	0
0.1	0.18	1.18	18
0.2	0.21	1.36	36
0.3	0.24	1.55	55
0.4	0.27	1.75	75
0.5	0.30	1.94	94

Table 2: Values of $nVar[F_n(0)]$, $nVar[F_n(.5)]$ and $nCov[F_n(0),F_n(.5)]$ for various values of correlation for a first order moving average Gaussian process.

It is worth pointing out that Table 2 is constructed from a very simple model, one which has correlations only between nearest "neighbors." The construction of a similar table for even a simple first order autoregressive (i.e., Markov) process would involve extensive calculations and numerical integration. However, Table 2 shows the essential features that the presence of even small correlations between only adjacent times produces large changes in the variance of $F_n(a)$ and the covariance between $F_n(a)$ and $F_n(b)$. Thus, a correlation of only 0.1 or -0.1 will produce a change of about 12% in the variances of $F_n(0)$ and $F_n(0.5)$ and about a 16% change in their covariance regardless of n.

Let us now proceed to generalize these observations.

Lemma 1.1: Let R be the bivariate Gaussian cumulative distribution function evaluated at (a,b):

$$R(\rho) = \int_{-\infty}^{a}\int_{-\infty}^{b} \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left[\frac{-1}{2(1-\rho^2)}(x^2+y^2-2\rho xy)\right] dx\, dy$$

where unit variance is assumed. Then R is a monotonically increasing function of ρ on the interval (-1.0,1.0). Furthermore, its derivative is given by:

$$\frac{dR}{d\rho} = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left[\frac{-1}{2(1-\rho^2)}(a^2+b^2-2\rho ab)\right].$$

Proof: Define:

$$P = P(c<X_1<a, d<X_2\leq b)$$

$$= \int_c^a \int_d^b f(x_1,x_2)\, dx_1\, dx_2$$

where f(.,.) is the bivariate Gaussian density with means 0, variances 1, and correlation ρ. We have by definition of the characteristic function of $f(x_1,x_2)$:

$$f(x_1,x_2) = \frac{1}{(2\pi)^2} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \exp(-it'x)\exp(-\frac{1}{2}t'Vt)\ dt_1\ \ dt_2$$

where V is the covariance matrix of the 2-vector $x=(x_1,x_2)$ and t is a 2-vector, $t=(t_1,t_2)$. Then after substituting, we get:

$$P = \int_c^a \int_d^b \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{1}{(2\pi)^2} \exp(-it'x)\exp(-1/2t'Vt)\ dt_1\ dt_2\ dx_1\ dx_2$$

Interchanging the order of integration yields:

$$P = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{1}{(2\ \pi\ )^2} \int_c^a \int_d^b \exp(-it'x)\ \exp(-1/2t'Vt)\ dx_1\ dx_2\ dt_1\ dt_2$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{1}{(2\ \pi\ )^2} \exp(-1/2t'Vt) \frac{[e^{-it_1c} - e^{-it_1a}]}{it_1} \frac{[e^{-it_2d} - e^{-it_2b}]}{it_2}\ dt_1dt_2$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{1}{t_1t_2(2\pi)^2}\ e^{-1/2t'Vt}\ \ [e^{-i(t_1c+t_2d)} - e^{-i(t_1c+t_2b)}$$

$$-e^{-i(t_1a+t_2d)} + e^{-i(t_1a+t_2b)}]\ dt_1\ dt_2$$

Now,

$$V = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$$

so,

$$t'Vt = t_1^2 = t_2^2 + 2\rho t_1t_2$$

Thus, taking derivatives inside the integral gives:

$$\frac{dP}{d\rho} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{1}{(2\pi)^2}\ \ e^{-1/2t'Vt}\ \ [e^{-i(t_1c+t_2d)} - e^{-i(t_1c+t_2b)}$$

$$- e^{-i(t_1a+t_2d)} + e^{-i(t_1a+t_2b)}]\ dt_1\ dt_2$$

But this integral is just the algebraic sum of four bivariate normal densities, that is:

$$\frac{dP}{d\rho} = f(c,d) + f(a,b) - f(c,b) - f(a,d)$$

Now if we take limits as $c \longrightarrow -\infty$ and $d \rightarrow -\infty$, then $P \rightarrow R(\rho)$ and so:

$$\frac{dR}{d\rho} = \frac{dP}{d\rho} = f(a,b) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left[\frac{-1}{2(1-\rho^2)} (a^2+b^2-2\rho ab)\right]$$

Clearly, $\frac{dR}{d\rho}$ is greater than 0 for any $\rho \in (-1,1)$ and for any a and b.

Q.E.D.

As a result of this lemma, we may write the indefinite integral:

$$I(a,b) = \int \frac{dR}{d\rho}\, d\rho$$

$$= \int \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left[\frac{-1}{2(1-\rho^2)}(a^2+b^2-2\rho ab)\right] d\rho$$

If $a=b=0$, we get:

$$I(0,0) = \int \frac{1}{2\pi\sqrt{1-\rho^2}}\, d\rho$$

which, upon integrating, yields Sheppard's theorem on median dichotomy (Kendall & Stuart, Vol. 1, [1969]).

We are now prepared to prove our central result.

Theorem 2: Let X(t) be a stationary Gaussian process defined for integer time with mean 0 and variance 1. Let ρ_i be the correlation between X(t) and X(t+i). Then, $Var[F_n(a)]$ is a monotonically increasing function of ρ_i for $\rho_i \in (-1,1)$. Furthermore, we have that:

$$\frac{\partial \, Var}{\partial \rho_i}[F_n(a)] = \frac{2}{n}(1-\frac{i}{n})\frac{1}{2\pi\sqrt{1-\rho_i^2}} \exp[\frac{-a^2}{1+\rho_i}]$$

Proof: We have:

$$Var[F_n(a)] = \frac{1}{n}F(a)\ [1-F(a)] + \frac{2}{n}\sum_{1}^{n}(1-\frac{i}{n})[P(X(0)\leq a, X(i)\leq a) - F(a)^2]$$

Let the variance (unity) and ρ_j be maintained constant for $j \neq i$. Then:

$$\frac{\partial \, Var}{\partial \rho_i}[F_n(a)] = \frac{2}{n}(1-\frac{i}{n})\frac{\partial P}{\partial \rho_i}(X(0)\leq a, X(i)\leq a)$$

By Lemma 1 and since $i\leq n$, we have:

$$\frac{\partial \, Var}{\partial \rho_i}[F_n(a)] = \frac{2}{n}(1-\frac{i}{n})\frac{1}{2\pi\sqrt{1-\rho^2}} \exp[\frac{-a^2}{1+\rho_i}]$$

This is non-negative for any a and for $\rho_i \in (-1,1)$.

Q.E.D.

We also have a monotonicity property for the covariance between $F_n(a)$ and $F_n(b)$:

Theorem 3: Let X(t) be a stationary Gaussian process defined for integer time with mean 0 and variance 1. Let ρ_i be the correlation between X(t) and X(t+i). Then $Cov[F_n(a), F_n(b)]$ is a monotonically increasing function of ρ_i for $\rho_i \in (-1,1)$. Furthermore, we have:

$$\frac{\partial \, Cov}{\partial \rho_i}[F_n(a),F_n(b)] = \frac{2}{n}(1-\frac{i}{n})\frac{1}{2\pi\sqrt{1-\rho_i^2}}\exp[\frac{-1}{2(1-\rho_i^2)}(a^2+b^2-2\rho_i ab)]$$

Proof: For ease of notation, let $C = Cov[F_n(a),F_n(b)]$. Let $a \leq b$, clearly without loss of generality. We have, by an earlier result:

$$C = \frac{1}{n}F(a)\,[1-F(b)] + \frac{1}{n}\sum_1^n (1-\frac{i}{n})\,[P(X(i)\leq a,X(0)\leq b)-F(a)F(b)]$$

$$+ \frac{1}{n}\sum_1^n (1-\frac{i}{n})\,[P(X(0)\leq a,X(i)\leq b)-F(a)F(b)]$$

If we maintain the variance (=1) and ρ_j constant for $j \neq i$, we get:

$$\frac{\partial C}{\partial \rho_i} = \frac{1}{n}(1-\frac{i}{n})\frac{\partial P}{\partial \rho_i}(X(i)\leq a,X(0)\leq b) + \frac{1}{n}(1-\frac{i}{n})\frac{\partial P}{\partial \rho_i}(X(0)\leq a,X(i)\leq b)$$

$$= \frac{2}{n}(1-\frac{i}{n})\frac{\partial P}{\partial \rho_i}(X(0)\leq a,X(i)\leq b)$$

This last step follows from stationarity ($\rho_i = \rho_{-i}$) and from the symmetry of the bivariate Gaussian density. By Lemma 1, this becomes:

$$\frac{\partial C}{\partial \rho_i} = \frac{2}{n}(1-\frac{i}{n})\frac{1}{2\pi\sqrt{1-\rho_i^2}}\exp[\frac{-1}{2(1-\rho_i^2)}(a^2+b^2-2\rho_i ab)]$$

The monotonicity now follows from Lemma 1.

Q.E.D.

It is tempting to immediately extrapolate this monotonicity property of $Var[F_n(a)]$ to $Var[f_n(a,b)]$. We have:

$$Var[f_n(a,b)] = Var[F_n(a)] + Var[F_n(b)] - 2Cov[F_n(a),F_n(b)]$$

After substituting the results of the previous two theorems, we get:

$$\frac{\partial \operatorname{Var}[f_n(a,b)]}{\partial \rho_i} = \frac{2}{n}\left(1-\frac{i}{n}\right)\frac{1}{2\pi\sqrt{1-\rho_i^2}}\left[\exp\left(\frac{-a^2}{1+\rho_i}\right)+\exp\left(\frac{-b^2}{1+\rho_i}\right)\right.$$

$$\left.-2\exp\left(\frac{-1}{2(1-\rho_i)}(a^2+b^2-2\rho_i ab)\right)\right]$$

Let G be the term in brackets. Now

$$a^2+b^2-2\rho ab = a^2-\rho a^2+b^2-\rho b^2+\rho a^2+\rho b^2-2\rho ab$$

$$= a^2(1-\rho)+b^2(1-\rho)+\rho(b-a)^2$$

So, G can be written as:

$$G = \exp\left(\frac{-a^2}{1+\rho}\right)+\exp\left(\frac{-b^2}{1+\rho}\right)-2\exp\left(\frac{-a^2}{2(1+\rho)}\right)\exp\left(\frac{-b^2}{2(1+\rho)}\right)\exp\left(\frac{-\rho(b-a)^2}{2(1-\rho^2)}\right)$$

where the subscript i has been omitted for simplicity. Now if $\rho \geq 0$, then:

$$\exp\left(\frac{-\rho(b-a)^2}{2(1-\rho^2)}\right) \leq 1$$

So,

$$G \geq \exp\left(\frac{-a^2}{1+\rho}\right)+\exp\left(\frac{-b^2}{1+\rho}\right)-2\exp\left(\frac{-a^2}{2(1+\rho)}\right)\exp\left(\frac{-b^2}{2(1+\rho)}\right)$$

$$= \left[\exp\left(\frac{-a^2}{2(1+\rho)}\right)-\exp\left(\frac{-b^2}{2(1+\rho)}\right)\right]^2$$

$$\geq 0$$

Thus, we have established the following corollary:

Corollary 3.1: Let $X(t)$ be a stationary Gaussian process defined for integer time with mean 0 and variance 1. Let ρ_i be the correlation between $X(t)$ and $X(t+i)$. Then $Var[f_n(a,b)]$ is a monotonically increasing function of ρ_i for $\rho_i \in [0,1]$.

It is worth emphasizing the difference between this corollary and the previous theorems. The distinction lies in the range of the correlation: $Var[f_n(a,b)]$ is monotonic only for non-negative correlations, while $Var[F_n(a)]$ is monotonic for all correlations.

4. DISCUSSION

We have presented a framework for probability estimation from stochastic processes that is a generalization of that from independent, identically distributed variables. Thus, the requirement of stationarity is an extension of that of being identically distributed. Similarly, the requirement of α-mixing is an extension of the notion of independence; indeed, α-mixing implies asymptotic independence and is thus a satisfying generalization. We have seen that, under the conditions of stationarity and α-mixing, the empirical distribution function is an unbiased and consistent estimator of the univariate cumulative distribution function. The same applies to the empirical amplitude histogram as an estimator of the theoretical amplitude histogram.

We have shown that the presence of correlations is manifested in the variance and covariance of the estimators. Specifically, these variances and covariances have been expressed as a sum of two terms: the first is simply that for independent variables, while the second is the correction due to dependence. By s simulation, we showed numerically that even small correlations may have a pronounced effect on the variance and covariance. A monotonicity property of the bivariate Gaussian distribution was used to characterize the effect of correlations for Gaussian processes. It was established that the variances and covariances are monotonically related to the correlations for Gaussian processes.

The convergence of the estimators as the sample size increases is assured by the consistency property. We may regard the estimator variance as a measure of the rate of convergence in sample size. Specifically, the variance gives the average (over the ensemble) rate of convergence in mean square. Thus, the monotonicity of the variance implies a monotonicity of the rate of convergence. For Gaussian processes, large positive correlations imply slow convergence and, consequently, poor estimates compared to the same sized samples of independent data. There is another aspect of the

monotonicity property which should be noted. This is that the variance and covariance of the empirical distribution function decreases as the correlations become more negative. Negative correlations are beneficial in the sense that they result in better estimates than the independent case. This property of negative correlations is not shared by the amplitude histogram.

Although our focus has been on probability estimation, our results have consequences for hypothesis testing with correlated data. This is especially true for hypothesis tests of univariate Gaussianity. Weiss (1978) has devised an empirical correction formula for the critical values of the Kolmogorov-Smirnov test for Gaussianity when there are correlations present. Since this test is based on a comparison of the empirical and Gaussian distribution functions, we would expect our results to be applicable. This modified Kolmogorov-Smirnov test exhibited very poor power in Weiss' simulations of data dominated by positive correlations, i.e., poor compared to the power of the test for independent variables. Weiss effectively could not distinguish between correlated data generated by an autoregression from Gaussian variables from that generated from uniform variables. Our monotonicity theorems suggest a possible explanation: the empirical distribution function of the Gaussian process cannot be determined as accurately when there are positive correlations dominating the data. It is this inherent inaccuracy of the measurement that may be responsible for this poor power.

REFERENCES

Abramowitz, M. and Stegun, I. A. (eds.), Handbook of Mathematical Functions, National Bureau of Standards, U.S. Department of Commerce, Washington, D.C., 1964.

Billingsley, P., Probability and Measure, Wiley, New York, 1979.

Box, G. E. P. and Jenkins, G. M., Time Series Analysis: Forecasting and Control, revised edition, Holden-Day, San Francisco, California, 1976.

Elul, R., "Gaussian Behavior of the Electroencephalogram: Changes During Performance of a Mental Task," Science, 164: 328-331, 1969.

Fuller, W. A., Introduction to Statistical Time Series, Wiley, New York, 1976.

Gasser, T., "Goodness-of-fit Tests for Correlated Data," Biometrika, 62: 563-570, 1975.

Kendall, M. G. and Stuart, A., The Advanced Theory of Statistics, Vol. 1, Hafner, New York, 1969.

Leadbetter, M. R. and Watson, G. S., "On the Estimation of the Probability Density," ASIA 264811, Alexandria, Virginia, 1961.

McEwen, J. and Anderson, G., "Modeling the Stationarity and Gaussianity of Spontaneous Electroencephalographic Activity," IEEE Trans. Biomed. Eng., BME-22: 361-369, 1975.

Oppenheim, A. V. and Schafer, R. W., Digital Signal Processing, Prentice-Hall, Englewood Cliffs, New Jersey, 1975.

Parzen, E., "On the Estimation of a Probability Density Function and Mode," Ann. Math. Stat., 33: 1065-1076, 1962a..

_______, Stochastic Processes, Holden-Day, San Francisco, California, 1962b.

Patankar, V. N., "The Goodness of Fit of Frequency Distributions Obtained from Stochastic Processes," Biometrika, 41: 450-462, 1954.

Persson, J., "Comments on Estimation and Tests of EEG Amplitude Distributions," Electroenceph. Clin. Neurophysiol., 37: 309-313, 1974.

Saunders, M. G., "Amplitude Probability Density Studies on Alpha and Alpha-like Patterns," Electroenceph. Clin. Neurophysiol., 15: 761-767, 1963.

Thrall, G. P., "An Analysis of Amplitude Probability Measurements," FDL-TDR-64-116, Research and Technology Division, AFSC, Wright-Patterson AFB, Ohio, 1965.

Wezman, E. J., "Nonparametric Probability Estimation: I. A Summary of Available Methods," Technometrics, 14: 533-546, 1972.

Weiss, M., "Modification of the Kolmogorov-Smirnov Statistic for Use with Correlated Data," J. Am. Stat. Assoc., 73: 872-875, 1978.